ESTIMATION OF MIXTURES OF PROBABILISTIC PCA WITH STOCHASTIC EM FOR THE 3D BIPLANAR RECONSTRUCTION OF SCOLIOTIC RIB CAGE

S. Benameur[abc],   M. Mignotte[ab],   F. Destrempes[b],   and J.A. De Guise[ac]

[a] Laboratoire de recherche en imagerie et orthopédie, Center hospitalier Universitaire de Montréal, Canada

[b] Laboratoire de vision et modélisation géométrique, DIRO, Université de Montréal, Canada

[c] École de technologie supérieure, Montréal, Canada

E-MAIL : BENAMEUS@IRO.UMONTREAL.CA

ABSTRACT

In this paper, we present a robust method for estimating the model parameters in a mixture of probabilistic principal component analyzers. This method is based on the Stochastic version of the Expectation Maximization (SEM) algorithm. Parameters of this mixture model are herein used to constrain the 3D reconstruction problem of scoliotic rib cage from a pair of planar and conventional calibrated radiographic images (postero-anterior with normal incidence (I_{PA}) and lateral (I_{LAT})). More precisely, the proposed PPCA mixture model is herein robustly exploited for dimensionality reduction and to get a set of probabilistic prior models associated to each detected class of pathological deformations observed on a representative training scoliotic rib cage population. By using an appropriate likelihood and for each considered class-conditional prior model, the proposed 3D reconstruction is stated as an energy function minimization problem, which is solved with a stochastic optimization algorithm. The optimal 3D reconstruction then corresponds to the class of deformation and parameters leading to the minimal energy. This 3D method of reconstruction has been successfully tested on several biplanar radiographic images, yielding very promising results.

1. INTRODUCTION

The 3D reconstruction problem of scoliotic rib cage from two projections has not been widely studied and only few references exist in the literature. Recently, Mouren [1] has proposed a method allowing to reconstruct the 3D geometry of the rib cage from two radiographic projections (postero-anterior and lateral) and a global *prior* knowledge on the geometrical structure of each rib. To find this geometrical knowledge while reducing the dimensionality of this problem, a Principal Component Analysis (PCA) is applied to a training scoliotic rib cage population. This method is interesting but may not be very accurate. First, it is widely supervised and highly operator-dependent ; it requires to manually identify and digitize a set of 60 points in the lateral view. Moreover, the PCA only defines a *linear* dimensionality reduction, which is a strong and not necessarily true assumption in this context.

In order to overcome this problem of supervision and improve the accuracy of the deformation model, we propose herein to use a mixture of Probabilistic PCA [2] (PPCA). In order to estimate the parameters of such a mixture model, the Expectation Maximization (EM) algorithm was already proposed [2]. Nevertheless, the initial parameter values have a significant impact on the convergence of this iterative procedure and on the quality of the final estimation (The EM converges to a *local* and not necessary *global* optimum estimate). In order to give the procedure more robust, we herein propose a stochastic version of the EM-PPCA relying on the Stochastic version of the EM (SEM) algorithm [3]. This SEM-PPCA is herein efficiently exploited for dimensionality reduction and to get a set of probabilistic prior models, associated to each detected class of pathological deformations, observed on a representative training scoliotic rib cage population. For each considered class-conditional prior model of pathological deformations, the proposed reconstruction method then consists in fitting the projections of an instance of the deformation model with the segmented contours of the corresponding rib cage on the two radiographic views. This matching problem leads to a set of K optimization problems (one for each detected class of pathological deformations) which is efficiently solved in our application with a stochastic optimization algorithm. The optimal 3D reconstruction corresponds to the solution, leading to the minimal energy, amongst these K optimization problems.

This paper is organized as follows. In Section 2 we present the basic concept of mixtures of PPCA. Section 3 is devoted to the SEM algorithm for estimating of the model parameters in a mixtures of PPCA. Sections 4 and 5 describe the proposed 3D reconstruction method and the related sequence of energy function to be optimized. In Section 6, we show some 3D reconstruction results.

2. PROBABILISTIC MODEL FOR DIMENSIONALITY REDUCTION

PCA is a well established model for dimensionality reduction. Nevertheless, one limiting disadvantage of this technique is the absence of an associated probability density model or generative model [2]. PPCA overcomes this problem.

2.1. Probabilistic PCA

Here, we recall the concept and formulation of PPCA. Let S and b be two random vectors related by [2],

$$S = \Phi b + \nu + \varepsilon,$$

where Φ is a $d \times m$ matrix that represents the principal subspace of S. The assumption is that $b \sim \mathcal{N}(0, I_m)$ as well as $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_d)$ are zero mean Gaussian distributed random vector (I_r denotes the $r \times r$ identity matrix and $\mathcal{N}()$ represents the normal distribution).

Consequently the random vector S is also normally distributed according to

$$p_S(s) = \mathcal{N}(s; \nu, \sigma^2 I_d + \Phi\Phi^T). \qquad (1)$$

The variable S represents the full data, whereas b represents the reduced data (the dimension m of b is signi£cantly smaller than the dimension d of S).

Let $[s_1, s_2, ..., s_n]^T$ be a i.i.d. observation sample issued from S, $\lambda_1, ..., \lambda_d$ be the eigenvalues of the covariance matrix, in decreasing order, of this sample, Λ_m be the diagonal matrix with entries $\lambda_1, ..., \lambda_m$, and U_m be the $d \times m$ matrix with columns equal to the corresponding eigenvectors, normalized so that they have euclidean norm equal to 1. From [2], a ML estimation of (ν, σ^2, Φ) is given by,

$$\nu = \bar{s} = \frac{1}{n} \sum_{i=1}^{n} s_i,$$

$$\sigma^2 = \frac{1}{d-m} \sum_{i=m+1}^{d} \lambda_i,$$

$$\Phi = U_m(\Lambda_m - \sigma^2 I_m)^{1/2}. \qquad (2)$$

The conditional probability distribution of S given b, is expressed by [2]

$$P_{S/b}(s/b) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}\|s - W - \bar{s}\|\right),$$

where $W = U_m(\Lambda_m - \sigma^2 I_m)^{1/2}b$. Hence, the ML reconstructed data point should be taken as

$$\hat{s} = U_m(\Lambda_m - \sigma^2 I_m)^{1/2}\hat{b} + \bar{s}.$$

2.2. Mixtures of PPCA

A Mixture of PPCA (MPPCA) is a distribution of the form[2]

$$P_S(s) = \sum_{c=1}^{K} \omega_c \mathcal{N}(s; \nu_c, \sigma_c^2 I_d + \Phi_c\Phi_c^T),$$

where K designates the number of components of the mixture and ω_c is the proportion of each component c. This mixture depends on a vector of parameters $\Gamma = (\nu_1, \sigma_1, \Phi_1, ..., \nu_K, \sigma_K, \Phi_K)$. In order to reduce data, the following strategies can be adopted ; we compute for each mixture component c, the reconstruction point $\bar{s}_c$ and take the one nearest to s the one that maximizes the distribution $P_S(s)$ (optimality in the ML sense).

3. ESTIMATION OF A MIXTURES OF PPCA

In order to estimate the parameter vector Γ, we resort to the SEM algorithm [3]. This iterative procedure requires initial parameters which can be given by a K-means clustering procedure [4] (in doing so, we assume, as £rst approximation, that the considered clusters are spherical with equal volumes). The obtained spherical partitions allow to obtain a rough estimation of the mixture parameters which are then used to initialize the SEM clustering and estimation procedure.

3.1. Stochastic EM Algorithm

Let K be an upper bound on the number of classes and $\xi \in]0, 1[$ be a chosen threshold. In the context of the estimation of a MPPCA, the SEM algorithm [5] can be outlined as follows, (the superscript denoting the iteration number).

Initialization
We initialize the SEM algorithm by the parameters estimated on the complete data obtained by K-means algorithm clustering [4] of S into K classes. For every observation s_i, a probability $P_{S/C}^{[0]}(s_i/c)$ of its belonging to the class c, $(c \in \{1, ..., K\})$ can be de£ned. An iteration of SEM consists of three steps.

Step S (Stochastic)
For each s_i, we select from the set of classes an element according to the distribution $[P_{C/S}^{[p]}(1/s_i, \Gamma^{[p]}), ..., P_{C/S}^{[p]}(K/s_i, \Gamma^{[p]})]$. This selection de£nes a partition $[\vartheta_1^{[p]}, ..., \vartheta_K^{[p]}]$ of the sample $S = [s_1, s_2, ..., s_n]^T$.

Step M (Maximization)
The SEM algorithm supposes that every s_i belonging to $\vartheta_c^{[p]}$ for each $c \in [1, K]$, is realized according to the distribution de£ned by $P_{S/C}(s_i/c, \Gamma)$ (Eq. (1)), the density corresponding to the class c. Let N be the number of rib cages in the training base. By denoting $\hat{\pi}_c^{[p+1]} = \text{Card}(\vartheta_c^{[p]})/N$, we can estimate $\Gamma^{[p+1]}$, the parameter of the mixture, in the ML sense (Eq. (2)). If $\hat{\pi}_c^{[p+1]} < \xi$, we eliminate the class c in the mixture.

Step E (Estimation)
For each s_i, we de£ne the next distribution $[P_{C/S}(1/s_i, \Gamma^{[p+1]}), ..., P_{C/S}(K/s_i, \Gamma^{[p+1]})]$ by the posterior distribution based on the current parameter $\Gamma^{[p+1]}$:

$$P_{C/S}(c/s_i, \Gamma^{[p+1]}) = \frac{\pi_c^{[p+1]} P_{S/C}(s_i/c, \Gamma^{[p+1]})}{\sum_{c=1}^{K} \pi_c^{[p+1]} P_{S/C}(s_i/c, \Gamma^{[p+1]})}.$$

Return to step S until the sequence Γ becomes steady.

4. MIXTURE OF STATISTICAL DEFORMABLE MODELS

In our application, the shape s of each rib cage of our training database is de£ned by a set of l control points or landmarks, which approximate the geometrical shape of the mid-lines of each rib. Each scoliotic rib cage, in the training population, is thus represented by the following $3l$ dimensional vector,

$$s = (x_1, y_1, z_1, ..., x_l, y_l, z_l)^T,$$

where $(x_i, y_i, z_i)^T$ are the Cartesian coordinates of the ith control point of the shape.

4.1. Training Phase

After aligning of the training shapes, considered as our observation sample, we estimate the parameters of a MPPCA with K classes. Each component of this MPPCA allows to reduce the dimensionality to m_c ($\ll 3l$), for each class c ($c \in [1, K]$) and allows to get a probabilistic deformable model with parameters ν_c, σ_c^2, and $\Phi_c = U_{m_c}(\Lambda_{m_c} - \sigma_c^2 I_{m_c})^{1/2}$. This gives us non-linear admissible and probabilistic deformations for the mean shape of each pathological deformations detected in our representative learning set.

4.2. Deformation Parameters

The globally deformed template for each class c is £nally de£ned by,

$$s_c \;=\; M(k,\alpha)[\bar{s}_c + \Phi_c b_c] + T, \qquad (3)$$

where,

- T and $M(k,\alpha)$ account for rigid deformations of the template, (T is a global translation vector, and $M(k,\alpha)$ performs a rotation and a scaling by k (in the x, y or z axes)).
- $\Phi_c = (\phi_1, \ldots, \phi_{m_c})$ is the matrix of the £rst m_c eigenvectors associated with the m_c largest eigenvalues and $\mathbf{b_c} = (b_{c,1}, \ldots, b_{c,m_c})^t$ is a vector containing the weights for these m_c deformation modes.

A global con£guration of the deformable rib cage template is thus described by $m_c + 9$ parameters corresponding to rigid transformations and m_c modal weights b_c.

4.3. Prior Energy Term

Let Θ be the random variable corresponding to the vector of deformations. We model the distribution of Θ by [6],

$$P_\Theta(s_c) \;=\; \mathcal{U}(T,k,\alpha)\exp\!\left(-\tfrac{1}{2}b^t(I_{m_c} - \sigma_c^2\Lambda_{m_c}^{-1})\,b\right).$$

Where $\mathcal{U}$ is the uniform distribution. By considering that

$$P_B(S_c) \;=\; \frac{1}{\zeta_p}\exp(-E_p(s_c)),$$

where ζ_p is a normalization constant, A prior energy term $E_p(s_c)$ can be deduced,

$$E_p(s_c) \;=\; \frac{1}{2}b^t(I_{m_c} - \sigma_c^2\Lambda_{m_c}^{-1})\,b. \qquad (4)$$

This energy term, which does not penalize af£ne transformations, will be used in our energy-based model to constrain our 3D reconstruction model by penalizing the deviation of the deformed template from the mean shape.

4.4. Likelihood Energy Term

As proposed in [7], the likelihood model is expressed by the following heuristic likelihood energy term,

$$E_l(s_c, I_{\mathrm{PA}}, I_{\mathrm{LAT}}) = -\frac{1}{n_{\mathrm{PA}}}\sum_{\Gamma_{\mathrm{PA}}}\Psi_{\mathrm{PA}}(x,y) - \frac{1}{n_{\mathrm{LAT}}}\sum_{\Gamma_{\mathrm{LAT}}}\Psi_{\mathrm{LAT}}(x,y),(5)$$

where the summation of the £rst and second term of E_l is overall the n_{PA} and n_{LAT} points of the external contour of the respectively I_{LAT} and I_{PA} perspective projections of the deformed template, belonging to the class c, on two pre-computed edge potential £elds of each radiographic image.

The edge potential £eld Ψ is computed from Canny edge [8] and de£ned as in [9] by,

$$\Psi(x,y) \;=\; \exp\!\left(-\frac{\sqrt{\xi_x^2 + \xi_y^2}}{\tau}\right)\Big|\cos(\gamma(x,y))\Big|, \qquad (6)$$

where $\xi = (\xi_x, \xi_y)$ is the displacement to the nearest edge point in the image, and τ is a smoothing factor which controls the degree of smoothness of this potential £eld. $\gamma(x,y)$ is the angle between the tangent of the nearest edge and the tangent direction of the contour at (x,y) [7].

5. OPTIMIZATION OF THE ENERGY FUNCTION

For each class c ($\in [1,\ldots,K]$), the unsupervised 3D reconstruction procedure is stated as a energy function minimization problem, namely,

$$E(s_c,\theta) \;=\; E_l(s_c, I_{\mathrm{PA}}, I_{\mathrm{LAT}}) + \beta E_p(s_c),$$

where E_l is the likelihood energy term, E_p is the prior energy term or the regularization term. β is a factor allowing to control the balance between the two energy components and the rigidity of the deformable template.

In order to minimize this complex energy functions [7], we resort to the E/S optimization algorithm, recently proposed by O. François in [10], already used for shape localization [6], non photorealistic rendering [11], and 3D reconstruction of the scoliotic spine [7].

Finally, we consider that the optimal 3D reconstruction corresponds to the solution, leading to the minimal energy, amongst the K possible optimization problems

$$E_{\mathrm{optimal}}(s_c,\theta) = \arg\min_c\Big\{\arg\min_\theta E(s_c,\theta)\Big\}.$$

6. EXPERIMENTAL RESULTS

In our application, the rib cages database consists of $N = 532$ (representative) scoliotic rib cages. Each shape is represented by 660 mid-line points and have the same number of ribs (11 right ribs and 11 left ribs whose the mid line is composed of 30 points). The training phase yields a reduced dimension for the non-linear deformations with less than 15% error. The reduced dimensions, associated to each different classes, are presented in Table 1. For the experiments, we have chosen $\beta = 0.02$ for the weighting factor penalizing the prior energy term with respect to likelihood energy and $K = 8$ number of classes. We used the Canny edge detector to estimate the edge maps which are then used for estimation of the edge *potential £elds* on the two radiographic views (used in the likelihood energy term). In our application, $\sigma = 1$, mask size is 5×5, and the lower and upper thresholds are given by the unsupervised estimation technique proposed in [2]. We £x the size of population to 100 and the number of iterations to 100. Fig. 2 shows an example of 3D reconstruction of scoliotic rib cage from two radiographic images (postero-anterior and lateral).

class	1	2	3	4	5	6	7	8
reduced dim.	23	20	21	23	16	15	18	17

Table 1. Reduced dimension for each detected class of the PPCA mixture.

7. CONCLUSION

In this paper, we have presented a robust method for estimating all of the model parameters in a mixture of probabilistic principal component analyzers. This method is based on the Stochastic version of the Expectation Maximization (SEM) algorithm.

In order to validate this estimation technique, we have also presented a new technique for the 3D reconstruction of scoliotic rib cage from two radiographic projections I_{PA} and I_{LAT}. This method ef£ciently exploits the estimation of a mixture of PPCA for both dimensionality reduction and to constrain the 3D reconstruction problem. The proposed 3D reconstruction problem is viewed
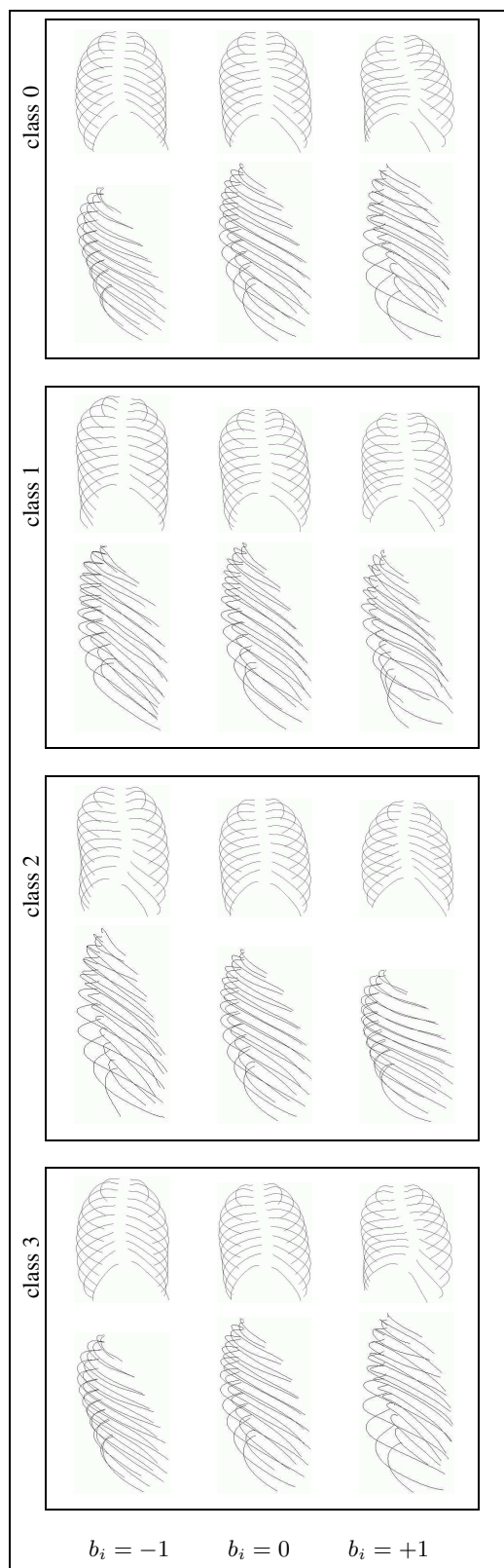
Fig. 1. Prior model of four £rst class. Two deformed shapes obtained by applying ± 1 of all deformation modes to the mean shape of rib cage and from the sagital and coronal views.
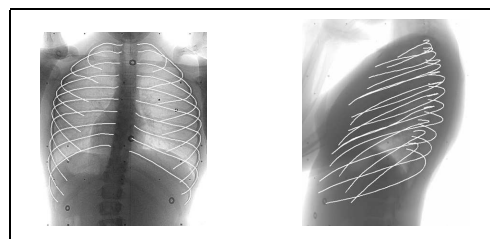

Fig. 2. Projections of reconstructed rib cage on the two radiographic views (i.e. postero-anterior and lateral views).

as a set of optimization problem, each one associated and constrain by a speci£c class of pathological deformation observed on a representative training scoliotic rib cage population and ef£ciently detected by the MPPCA method.

8. REFERENCES

[1] S. Mouren. Reconstruction 3D biplanaire de cages thoraciques scoliotiques à l'aide de modèles déformables de côtes. Master's thesis, Department of Mechanical Engineering, École Polytechnique at Montréal, Canada, 2003.

[2] M. E. Tipping and C. M. Bishop. Mixtures of probabilistic principal component analysers. *Neural Computation*, 11(2) :443–482, 1999.

[3] G. Celeux and J. Diebolt. L'algorithme SEM : un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités. *Revue de Statistique Appliquée*, 34(2) :35–52, 1986.

[4] J.A. Hartigan and M.A. Wong. A k-means clustering algorithm. *Applied Statistics*, 28 :100–108, 1979.

[5] P. Masson and W. Pieczynski. SEM algorithm and the unsupervised segmentation of satellite data. *IEEE Transactions on Geoscience Remote Sensing*, 31(3) :618–633, 1993.

[6] F. Destrempes and M. Mignotte. Unsupervised localization of shapes using statistical models. In *4th IASTED International Conference on Signal and Image Processing*, pages 60–65, Kauai Marriott, Hawaii, USA, August 2002.

[7] S. Benameur, M. Mignotte, S. Parent, H. Labelle, W. Skalli, and J. De Guise. A hierarchical statistical modeling approach for the unsupervised 3D reconstruction of the scoliotic spine. In *10th IEEE International Conference on Image Processing*, volume 1, pages 561–564, Barcelona, Spain, Sept. 2003.

[8] J. F. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6) :679–697, 1986.

[9] A. K. Jain, Y. Zhong, and S. Lakshmanan. Object matching using deformable templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(3) :267–278, 1996.

[10] F. Olivier. Global optimization with exploration/selection algorithms and simulated annealing. *The Annals of Applied Probability*, 12 :248–271, 2002.

[11] M. Mignotte. Unsupervised statistical sketching for non-photorealistic rendering models. In *10th IEEE International Conference on Image Processing*, volume 3, pages 573–577, Barcelona, Spain, Sept. 2003.